Serie 13

1) Zweiter t-Test

Ein Limonadenhersteller behauptet, dass eine Flasche seines Getränks nicht mehr als 18 Gramm Traubenzucker enthält. Eine Verbraucherorganisation nimmt eine Stichprobe von $n = 35$ Flaschen, und stellt einen durchschnittlichen Anteil von $\bar{x} = 18.67$ Gramm pro Flasche fest, bei einer geschätzten Standardabweichung von $\hat{\sigma}_x = 2$ Gramm. Zusätzlich gehen wir davon aus, dass für den Zuckergehalt der i-ten Flasche, $X_i \sim \mathcal{N}(\mu, \sigma_x^2)$ gilt.

(a) Testen Sie $H_0 : \mu = 18$ gegen die Alternative $H_A : \mu \neq 18$ mit einem t-Test.

(b) Berechnen Sie den P-Wert von dem obigen Test.
Dazu können Sie Mathematica verwenden, mit dem Kommando

```
CDF[StudentTDistribution[df],a]
```

um die Wahrscheinlichkeit $P[X \leq a]$ zu berechnen für eine t-verteilte Zufallsvariable X mit *df* Freiheitsgrade.
Zusätzlich kann man das Kommando `N[b]` verwenden, um den numerischen Wert von einem symbolischen Ausdruck b zu bestimmen.

(c) Bestimmen Sie noch ein 95%-Vertrauensintervall für den erwarteten Zuckergehalt μ.

2) Fortsetzung von Aufgabe 1: Mittelwertdifferenzen testen

Der Hersteller verspricht den Zuckergehalt zu senken. Ein Jahr später nimmt die Verbraucherorganisation eine weitere Stichprobe. Um die Probleme bei der Wahl des Tests zu vermeiden aber diesmal mit einer hohen Stichprobenanzahl von $n_2 = 500$. Es wird ein Mittelwert von $\bar{x}_2 = 17.9$ bei einer empirischen Standardabweichung von $\hat{\sigma}_2 = 3$ Gramm gemessen. Testen Sie $H_0 : \mu_1 = \mu_2$ gegen die Alternative $H_A : \mu_1 \neq \mu_2$ auf dem 5%-Niveau mit einem t-Test. Die Daten $n_1 = 35$, $\bar{x}_1 = 18.67$ und $\hat{\sigma}_1 = 2$ stammen dabei aus Aufgabe 2.

3) Output entziffern

Betrachte den folgenden Mathematica Output

```
Daten42 = {{2, -1, 0, 3}, {1, 0, 0, 2}, {2, 1, 1, 4},
  {1, 2, 1, 1}, {2, 3, 0, 0}}

{{2, -1, 0, 3}, {1, 0, 0, 2},
 {2, 1, 1, 4}, {1, 2, 1, 1}, {2, 3, 0, 0}}

M42 = LinearModelFit[Daten42, {x1, x2, x3}, {x1, x2, x3}]

FittedModel[ 0.375 + «19» x1 - «19» x2 + 1.75 x3 ]

M42["BestFit"](* Regressionsgleichung bestimmen *)

0.375 + 1.125 x1 - 0.875 x2 + 1.75 x3

M42["FitResiduals"](* Residuen bestimmen *)

{-0.5, 0.5, 0.5, -0.5, 1.77636×10^-15}

M42["ParameterTable"](* Die Parametertabelle *)
```

	Estimate	Standard Error	t Statistic	P-Value
1	0.375	1.65359	0.226779	0.858029
x1	1.125	0.927025	1.21356	0.43877
x2	−0.875	0.330719	−2.64575	0.230053
x3	1.75	0.968246	1.80739	0.321722

Darin wurde eine lineare Regression mit den gegebenen Daten in Mathematica ausgeführt. Lies Dir die Mathematica-Hilfe zum Befehl `LinearModelFit` durch, beantworte dann anhand des Outputs die folgenden Fragen:

(a) Wie lautet die von Mathematica gefundene Regressionsgleichung?

(b) Welche Residuen hat diese Gleichung für die Daten? Welcher Parameter ist am besten geschätzt?

(c) Welchen Y-Wert sagt das Modell für $(x^{(1)}, x^{(2)}, x^{(3)}) = (2, 3, 4)$ voraus?

(d) Die Koeffizienten $\hat{\beta}_j$ von der linearen Regression aus dem Output sind wie alle Schätzer unsicher. Bestimme ein Intervall um $\hat{\beta}_0$, das 95% der Wahrscheinlichkeitsmasse (bzgl. der $N(0,1)$-Verteilung) enthählt. Verwende bei der Standardisierung die in der Parametertabelle angegebene Standardabweichung.

4) Lineare Regression

Bestimme eine Regressionsebene $y = \beta^{(0)} + \beta^{(1)} x^{(1)} + \beta^{(2)} x^{(2)}$ für die Datenwerte

Stichprobe	1	2	3	4	5
Y	0	0	1	1	0
$X^{(1)}$	2	1	2	1	2
$X^{(2)}$	-1	0	1	2	3

Bestimme die Residuen dazu. Verwende für die Rechnung Mathematica, die entsprechenden Befehle finden sich in dem Mathematica Output von Aufgabe 1. Die Datentupel werden in der Form $(x^{(1)}, \ldots, x^{(n)}, y)$ definiert. Löse dann die Aufgabe per Hand und vergleiche beide Resultate.